

## 구간 중도절단 자료에서 Cox 비례위험모형의 추정방법 비교

박성환<sup>1)</sup>, 김진흠<sup>2)</sup>, 남정모<sup>3)</sup>

<sup>1)</sup>연세대학교 의과대학 의학통계학과, <sup>2)</sup>수원대학교 통계정보학과

<sup>3)</sup>연세대학교 의과대학 예방의학교실

### Comparison of Estimation Methods of Cox Proportional Hazards Model with Interval-censored Data

Sung Hwan Park<sup>1)</sup>, Jinheum Kim<sup>2)</sup>, Chung Mo Nam<sup>3)</sup>

<sup>1)</sup>Department of Biostatistics, Yonsei University College of Medicine

<sup>2)</sup>Department of Applied Statistics, University of Suwon

<sup>3)</sup>Department of Preventive Medicine, Yonsei University College of Medicine

#### Abstract

**Objectives:** Our purpose was to compare the methods for the regression coefficients estimation in a Cox proportional hazards model with interval-censored data.

**Methods:** The methods included the mid-point and right-endpoint imputation methods and 'intcox' method implemented following the algorithm proposed by Pan (1999, Journal of Computational and Graphical Statistics). Their performance was evaluated based on the estimated bias and mean squared error and the empirical coverage rate for the regression coefficients. Three methods were also discussed with real data analysis.

**Results:** The 'intcox' method had the better performance than the other imputation methods. In particular a 95% coverage rate of the 'intcox' method was very close to 0.95, while the imputation methods were pretty less than 0.95. As a right censoring rate decreases, all the methods underestimated the true value, but the 'intcox' method seemed to be more stable than the others owing to smaller bias and mean squared error. With analyzing the breast cosmetics data, the effect size of treatment based on the 'intcox' method was the largest among three methods, but there was no difference in statistical significance among them.

**Conclusions:** In many clinical studies the imputation methods were often used for dealing with the interval-censored data because they can be easily implemented using commercial softwares. However they may entail biased results as shown in simulation results such as low coverage rate. We recommend the 'intcox' method or multiple imputation methods rather than the single imputation methods.

**Key words:** Interval censoring, Cox proportional hazards model, 'intcox' method, Single imputation

[Submitted: 2015년 09월 17일, Revised: 2015년 10월 12일, Accepted: 2015년 10월 19일]

Corresponding author: Chung Mo Nam, PhD  
50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea  
Tel: +82-2-2228-1871  
E-mail: CMNAM@yuhs.ac

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A2056869) (Jinheum Kim).

## 1. 서론

생존분석(survival analysis)은 연구 시작시점부터 관심 있는 사건이 발생할 때까지의 시간을 다루는 임상연구에서 많이 사용되는 분석방법이다. 그러나 관심 있는 사건의 발생 시간이 구간으로 관측되는 경우가 대부분이다. 예를 들어 주기적으로 관측이 이루어지는 연구의 경우, 관심 사건이 관측주기 사이의 어느 시점에서 발생한다면 실제 사건발생 시간은 알 수 없고 대신 관심 사건의 발생 시간 전, 후 관측 시점에 대한 정보만 알 수 있다. 최근 임상에서는 이처럼 관심 사건의 발생 시간이 구간으로 관측되는 경우 즉, 구간 중도절단(interval-censored)된 경우 구간의 중간 값으로 대체하여 분석하거나 (중간값 대체법, mid-point imputation) 관측된 값의 우측 값으로 대체하여 분석하는 (우측값 대체법, right-endpoint imputation) 경우가 대부분이다. 하지만 이와 같은 분석방법은 편향된 결과를 초래할 수 있다. 따라서 이와 같은 문제를 해결하기 위해 다양한 방법들이 제안되었다.

Turnbull [1]은 자기일치성(self consistency)을 만족하는 비모수적 생존함수 추정량을 제안하였고, Groeneboom and Wellner [2]는 표본의 크기가 클 때 Turnbull [1]이 제안한 추정량의 수렴속도가 느린 단점을 보완하여 iterative convex minorant (ICM) 알고리즘을 제안하였다. Wellner and Zhan [3]은 EM 알고리즘과 ICM 알고리즘을 결합한 하이브리드 형태의 EM-ICM을 제안하였다. 또한 모형에 공변량을 포함시켜 공변량이 관심 사건의 위험률(hazard rate)에 미치는 영향을 살펴보기 위해 Finkelstein [4]은 Cox 비례위험모형(Cox proportional hazards model)에 뉴턴-랩슨(Newton-Raphson) 알고리즘을 적용하여 회귀계수(regression coefficients)를 추정하였다. Pan [5]은 기저위험함수(baseline hazard function)에 대해 조각지수분포(piecewise exponential distribution)를 가정하고 Groeneboom and Wellner [2]의 ICM

알고리즘을 적용하여 회귀계수를 추정하였다.

본 논문에서는 구간 중도절단된 자료를 분석하는 방법으로 Pan [5]의 알고리즘을 구현한 'intcox' 방법과 임상에서 많이 쓰이는 중간값 대체법과 우측값 대체법에 의한 회귀계수의 신뢰구간추정에 대해 살펴보고, 모의실험과 구간 중도절단된 자료의 예로 많이 사용되는 Finkelstein and Wolfe [6]의 유방암 자료 분석을 통해 세 가지 추정방법을 비교하고자 한다.

## 2. 연구 방법

### 1) 구간 중도절단 자료와 Cox 비례위험모형

본 논문에서 사용하는 자료를  $D=\{(A_i, x_i), i=1, \dots, n\}$ 로 정의한다. 여기서,  $n$ 은  $p$ -차원 공변량 벡터이고,  $A_i=(L_i, R_i)$ 이다. 특히, 생존시간  $T_i$ 가 좌측 중도절단(left-censored)되면  $L_i=0$ 이고, 정확히 관측되면  $L_i=R_i$ , 우측 중도절단(right-censored)되면  $R_i=\infty$ 이다. 자료  $D$ 에 대한 우도함수(likelihood function)는 다음과 같다.

$$L = \prod_{i=1}^n [S(L_i|x_i) - S(R_i|x_i)] \quad (1)$$

여기서,  $S(t|x) = \Pr(T > t | X = x)$ 이다. Turnbull [1]과 Lindsey and Ryan [7]이 제안한 방법에 따라  $A_i (i=1, \dots, n)$ 로부터 동등집합(equivalent class)을  $0 = t_0 < t_1 < \dots < t_m = \infty$ 로 정하면, 식 (1)은 다음과 같이 표현된다.

$$L = \prod_{i=1}^n \sum_{j=1}^m \alpha_{ij} [S(t_{j-1}|x_i) - S(t_j|x_i)].$$

여기서,  $(t_{j-1}, t_j] \subset A_i$ 이면  $\alpha_{ij}=1$ 로 정의하고, 그렇지 않으면  $\alpha_{ij}=0$ 으로 정의한다. 따라서 Cox 비례위험모형에 기초한 우도함수는 다음과 같이 표현된다.

$$L(S_0(t), \beta | D) = \prod_{i=1}^n \sum_{j=1}^m \alpha_{ij} [S_0(t_{j-1})^{\exp(\beta'x_i)} - S_0(t_j)^{\exp(\beta'x_i)}].$$

여기서,  $S_0(t)$ 는 기저생존함수(baseline survival-

al function)이고,  $\beta$ 는  $p$ -차원 회귀계수 벡터이다. 우측 중도절단된 자료에서는 부분우도 원리(principle of partial likelihood)를 적용하여 기저위험함수를 장애모수( nuisance parameter)로 간주할 수 있기 때문에  $\beta$ 에 대한 추론이 비교적 간단하다. 그러나 구간 중도절단된 자료에서는 회귀계수와 기저생존함수를 동시에 추정해야 하는 어려움이 있다.

Finkelstein [4]은 뉴턴-랩슨 알고리즘을 적용하여  $\beta$ 와  $S_0$ 에 대한 최대우도추정량(maximum likelihood estimate, MLE)을 제안하였다. 하지만 이 방법은 헤시안 행렬(Hessian matrix)의 역행렬을 계산해야 하는데 이로 인해 추정량이 수렴하지 않거나 수렴하더라도 수렴속도가 느린 단점을 가지고 있다. 그래서 Pan [5]은 기저위험함수에 대해 조각지수분포를 가정하고 Groeneboom and Wellner [2]의 ICM 알고리즘을 적용하여 추정하였다. ICM 알고리즘에서는  $\beta$ 와  $S_0$ 에 대한 추론을 위해 로그우도함수(log-likelihood function)를  $\beta$ 와  $S_0$ 에 대해 각각 두 번씩 편미분하여 얻은 다음 대각행렬

$$\begin{bmatrix} G_1(S_0, \beta) & 0 \\ 0 & G_2(S_0, \beta) \end{bmatrix}$$

에 음수를 취한 행렬로 뉴턴-랩슨 알고리즘의 헤시안 행렬을 대체하였다. 이 알고리즘은 뉴턴-랩슨 알고리즘보다 비교적 간단하며 수렴속도가 빠른 장점이 있다.

## 2) 분석방법

### (1) 대체법을 이용한 방법

결측치가 있는 경우 그 자료를 처리하는 방법으로 Rubin [8]이 제안한 대체법이 널리 사용되고 있다. 대체법은 크게 단일 대체법(single imputation)과 다중 대체법(multiple imputation)으로 구분할 수 있다.

본 논문에서는 구간 중도절단된 자료를 단일 대체법인 중간값 대체법과 우측값 대체법을 적용하여 우측 중도절단된 형태의 자료로 변환한 후

회귀계수를 추정하고자 한다. 구간  $(L, R]$  (단,  $R < \infty$ )에서 구간 중도절단된 자료에 대해, 중간값 대체법은 정확하게 관측되지 않은 관심사건의 발생시점을 두 시점의 중간 값인  $(L+R)/2$ 로 대체하는 방법이고, 우측값 대체법은 두 시점의 우측 값인  $R$ 로 대체하는 방법이다.  $R = \infty$ 이면 우측 중도절단된 자료 그대로 다룬다.

구간 중도절단된 자료를 대체법을 써서 우측 중도절단된 자료로 바꾼 후, 그 자료에 Cox [9]의 부분우도 원리를 적용하여 회귀계수를 추정하였고 또한 그에 대한 Wald 신뢰구간을 추정하였다. 이때 통계소프트웨어 R에 내장된 'coxph' 함수를 사용하였다.

### (2) 'intcox' 방법을 이용한 추정

본 논문에서는 Pan [5]이 제안한 알고리즘을 사용하여 Henschel et al. [10]이 구현한 통계소프트웨어 R에 내장된 'intcox' 함수를 사용하여 회귀계수를 추정하였다. 하지만 'intcox' 함수는 회귀계수  $\beta$ 와 기저생존함수  $S_0$ 에 대한 MLE만 제공하고 추정량의 표준오차는 제공하고 있지 않다. 따라서 본 논문에서는 Henschel et al. [10], Kim [11]이 제안한 붓스트랩 방법을 적용하여 추정량의 표준오차를 구하였다.

표현을 간편하게 하기 위해 공변량이 1개 즉,  $p = 1$ 이라고 가정하자. 주어진 한 모의실험 자료에서  $B$ 번 반복적으로 붓스트랩을 수행하여 뽑은 표본집합을 각각  $D^{(1)} = \{(A_i^{(1)}, x_i^{(1)}), i = 1, \dots, n\}, \dots, D^{(B)} = \{(A_i^{(B)}, x_i^{(B)}), i = 1, \dots, n\}$  라고 하자. 'intcox' 방법을 각 표본집합에 적용하여 얻은 회귀계수의 추정 값을 각각  $\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(B)}$  라고 할 때, 회귀계수  $\beta$ 에 대한  $100 \times (1 - \alpha)\%$  신뢰구간은  $(\hat{\beta}_{(B \times \alpha/2)}, \hat{\beta}_{(B \times (1 - \alpha/2))})$ 으로 정의한다. 여기서  $\hat{\beta}_{(1)} \leq \dots \leq \hat{\beta}_{(B)}$ 는  $\hat{\beta}^{(b)}$  ( $b = 1, \dots, B$ )를 크기 순서로 나열한 것이다.

## 3) 모의실험방법 및 분석자료

본 논문의 모의실험에서 사용될 자료를 생성하기에 앞서 먼저 기호를 소개하면,  $K$ 는 첫 번째 방문 이후 방문한 총 횟수를 나타내고(즉, 총 방문 횟수= $(K+1)$ ), 'len'은 두 방문시점 사이의 간격으로 즉, 방문주기를 나타낸다. 본 모의실험에서는  $K$ 를 3과 5로 설정하고, 방문주기 len은 일양분포(uniform distribution)  $U(1, 3, 1.7)$ 에서 생성하였다. 구간 중도절단된 모의실험 자료는 Pan [12]이 제안한 방법을 적용하여 생성하였으며,  $i(i=1, \dots, n)$  번째 개체의 구간 중도절단된 자료를 생성하는 절차는 다음과 같다.

① 2차원 공변량 벡터  $(X_{i1}, X_{i2})'$ 는 평균벡터가  $\mu = (0, 0)'$ 이고, 공분산 행렬이  $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ 인 이변량 정규분포로부터 생성하였다. 여기서,  $\rho$ 는 두 변수,  $X_1, X_2$ 의 상관정도를 나타내어  $\rho$ 는 0, 0.1, 0.5로 설정하였다.

② 실제 생존시간(true survival time)  $T_i$ 는 다음 Cox 비례위험모형으로부터 생성하였다.

$$\lambda(t|X_{i1}, X_{i2}) = \lambda_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2}).$$

여기서,  $\beta_1 = \beta_2 = 1$ 로 설정하였다. 기저위험함수는 형태모수(shape parameter)가 0.1인 와이블 분포(Weibull distribution)로 설정하였으며, 척도모수(scale parameter)는 생존시간의 평균이 3이 되도록 설정하였다.

③ 첫 번째 방문시점  $Y_{i1}$ 는  $U(0, 1)$ 에서 생성하였으며,  $k(k=2, \dots, K+1)$  번째 방문시점은  $Y_{ik} = Y_{i1} + (k-1) \times \text{len}_i$ 로 정의하였다. 구간  $(0, \infty]$ 를  $(K+2)$ 개 구간  $(0, Y_{i1}], \dots, (Y_{i, K+1}, \infty]$ 으로 나누었다. 만일  $T_i$ 가 구간  $(Y_{il}, Y_{i, l+1}] (l=0, 1, \dots, K+1)$ 에 속하면, 구간 중도절단된 자료( $L_i, R_i$ )를  $(Y_{il}, Y_{i, l+1}]$ 로 정의하였다.

한편 본 논문의 실제자료 분석에 사용된 자료는 1976년부터 1980년까지 미국 보스턴에 있는 방사선 치료 센터에서 유방암 환자로부터 수집된 자료이다 [6]. 94명의 유방암 환자를 대상으로 46명의 환자에게는 방사능 치료만, 48명의 환

자에게는 방사능 치료와 화학요법을 병행하여 치료하였다. 이 연구에서는 두 치료법이 유방함몰(cosmetic deterioration)에 미치는 영향을 연구하고자 하였다. 모든 환자들은 4주에서 6주마다 병원에 내원하여 유방함몰여부에 대한 진단을 받았으며, 생존시간은 수술 후 유방함몰이 나타날 때까지의 시간으로 정의하였다. 연구기간 동안 유방함몰이 발생하지 않은 환자는 우측 중도절단된 것으로 정의하였다. 그 외 환자들의 정확한 유방함몰 발생시점은 관측할 수 없었으며 대신, 유방함몰로 진단되지 않은 마지막 병원 방문시점과 유방함몰로 진단된 병원 방문시점 사이의 어느 시점에서 유방함몰이 발생했다고 정의하였다.

### 3. 연구 결과

#### 1) 모의실험 연구 결과

Table 1은 표본크기  $n=200$ 이고 붓스트랩 횟수  $B=1000$ 인 경우의 모의실험 결과이다. 우측 중도절단된 비율(right censoring percentage, RC(%))을 살펴보면  $\rho$ 의 값에 관계없이  $K=3$ 인 경우에 비해  $K=5$ 인 경우에 낮았다. 이러한 결과는 첫 번째 방문시점의 분포와 방문간격의 분포는 동일하지만 방문횟수가 늘어나면 ( $K=3 \rightarrow 5$ ) 우측 중도절단된 비율이 감소할 수밖에 없기 때문이다.  $\rho=0$ 이고  $K=3$ 인 경우에 추정량의 편향(Bias)을 보면, 단일 대체법 모두 실제 값보다 과소 추정하는 경향을 보였는데 우측값 대체법이 중간값 대체법보다 다소 더 편향된 결과를 보였다. 반면 'intcox' 방법은 단일 대체법에 비해 실제 값에 더 가까운 결과를 보였다. 평균제곱오차(mean squared error, MSE)는 추정된 값들의 편차와 Bias를 함께 고려하기 위한 척도로 'intcox' 방법이 단일 대체법보다 다소 작은 것으로 나타났다. 95% 신뢰구간 포함률(coverage rate, CR)을 보면, 단일 대체법은 0.95보다 매우 작았지만 'intcox' 방법은 0.95에 가까운

결과를 보였다.  $\rho=0$ 일 때  $K$ 가 3에서 5로 증가하면, 'intcox' 방법에서는,  $\beta_1$ ,  $\beta_2$ 의 Bias가 다소 증가하였지만 MSE는 변동이 없었고 CR은 다소 증가하였다. 단일 대체법에서는  $\beta_1$ 의 Bias와 MSE는 증가하고 CR은 다소 감소했지만  $\beta_2$ 의 결과에는 변동이 없었다. 한편 두 공변량의 상관정도  $\rho$ 가 0에서 0.5로 늘어나더라도 'intcox'의 방법은 별다른 변동이 없었지만 단일 대체법은 Bias와 MSE가 모두 증가하였고 CR은 0.95보다 더 작아졌다.

요법의 복합 치료를 같이 받았다면  $x=1$ 이다. 따라서 공변량  $x$ 를 가진 생존함수에 대한 추정 식은 다음과 같다.

$$\hat{S}(t|x) = \hat{S}_0(t)^{\exp(\hat{\beta}x)}.$$

Table 2는 'intcox' 방법, 중간값 대체법, 우측값 대체법을 써서 유방암 자료를 분석한 결과이다. 방사능 치료만을 받은 그룹에 비해 방사능 치료와 화학요법의 복합 치료를 받은 그룹이 유방암몰의 위험률이 'intcox' 방법의 경우는  $2.174(=e^{0.777})$

**Table 1.** Estimated bias (Bias), mean squared error (MSE), and 95% coverage rate (CR) of the regression parameters,  $\beta_1$  and  $\beta_2$ , based on 1000 replications with true value of  $\beta_1=1$  and  $\beta_2=1$  according to three methods ('intcox', single imputation by mid-point, and single imputation by right-endpoint)

$\rho^1$	$K^2$	RC(%) <sup>3</sup>		Method								
				'intcox'			Single imputation by mid-point			Single imputation by right-endpoint		
				Bias	MSE	CR	Bias	MSE	CR	Bias	MSE	CR
0	3	22.0	$\beta_1$	0.022	0.012	0.935	-0.083	0.016	0.846	-0.086	0.017	0.849
			$\beta_2$	0.016	0.012	0.930	-0.089	0.018	0.820	-0.091	0.019	0.811
	5	9.3	$\beta_1$	-0.028	0.011	0.938	-0.09	0.018	0.803	-0.093	0.018	0.788
			$\beta_2$	-0.024	0.011	0.943	-0.084	0.016	0.834	-0.086	0.017	0.819
0.1	3	22.6	$\beta_1$	0.018	0.013	0.932	-0.094	0.019	0.809	-0.096	0.019	0.801
			$\beta_2$	0.013	0.013	0.928	-0.099	0.020	0.789	-0.102	0.021	0.785
	5	9.9	$\beta_1$	-0.032	0.010	0.947	-0.093	0.017	0.807	-0.096	0.018	0.802
			$\beta_2$	-0.028	0.011	0.942	-0.091	0.018	0.806	-0.093	0.019	0.787
0.5	3	24.7	$\beta_1$	0.014	0.017	0.923	-0.123	0.029	0.756	-0.126	0.030	0.743
			$\beta_2$	0.019	0.016	0.924	-0.119	0.028	0.779	-0.123	0.029	0.773
	5	12.6	$\beta_1$	-0.029	0.014	0.942	-0.109	0.024	0.790	-0.112	0.025	0.768
			$\beta_2$	-0.025	0.013	0.946	-0.108	0.023	0.786	-0.112	0.024	0.765

<sup>1</sup>Correlation coefficient between two covariates.

<sup>2</sup>Total number of visits scheduled after the first visit.

<sup>3</sup>Right censoring percentage.

## 2) 실제자료 연구 결과

본 연구에 사용된 자료는 유방암 환자를 대상으로 두 치료 방법이 유방암몰까지 시간에 미치는 영향을 연구하기 위한 자료이다. 공변량  $x$ 는 방사능 치료만 받았다면  $x=0$ 이고, 방사능 치료와 화학

배, 중간값 대체법의 경우는  $1.973(=e^{0.680})$  배, 우측값 대체법의 경우는  $1.977(=e^{0.681})$  배로 높았다. 단일 대체법은  $\hat{\beta}$ 와  $\beta$ 에 대한 신뢰구간이 모두 유사하였다. 'intcox' 방법은 단일 대체법보다  $\hat{\beta}$ 는 다소 크고,  $\beta$ 에 대한 신뢰구간의 폭도 다소 넓었다.



**Table 2.** Estimated regression parameter ( $\hat{\beta}$ ) and 95% bootstrap or asymptotic confidence interval according to three methods ('intcox', single imputation by mid-point, and single imputation by right-endpoint)

Method					
'intcox'		Single imputation by mid-point		Single imputation by right-endpoint	
$\hat{\beta}$	Bootstrap CI	$\hat{\beta}$	Asymptotic CI	$\hat{\beta}$	Asymptotic CI
0.777	(0.23, 1.45)	0.680	(0.13, 1.23)	0.681	(0.14, 1.23)

또한 세 가지 방법 모두 95% 신뢰구간 안에 0이 포함되지 않기 때문에 방사능 치료만 받은 그룹에 비해 방사능 치료와 화학요법의 복합치료를 받은 그룹이 유방암의 위험이 유의수준 5%에서 유의하게 높은 것을 알 수 있다.

#### 4. 고찰 및 결론

본 논문에서는 Cox 비례위험모형의 회귀계수 추정에서 구간 중도절단된 자료를 다루는 방법으로 쓰이는 'intcox' 방법과 중간값 대체법, 우측값 대체법을 모의실험을 통해 서로 비교하였고, 유방암 자료에 적용하여 그 결과를 비교, 해석하였다. 모의실험 결과에 의하면 단일 대체법은 비슷한 결과를 보였고, 'intcox' 방법은 단일 대체법에 비해 실제 값에 조금 더 가깝게 추정하였으며(편향의 크기가 작음), 평균제곱오차 또한 단일 대체법에 비해 작은 것으로 나타났다. 'intcox' 방법은 95% 신뢰구간 포함률이 0.95와 거의 같았지만 단일 대체법은 0.95보다 매우 작았다. 한편 구간 중도절단된 비율이 증가(우측 중도절단된 비율이 감소)함에 따라 세 방법 모두 과소 추정하는 경향이 있었지만 'intcox' 방법은 단일 대체법에 비해 더 안정적으로 추정한다는 것을 알 수 있었다. 공변량들의 상관정도가 늘어나더라도 'intcox' 방법은 변동이 없었지만 단일 대체법은 편향과 평균제곱오

차가 모두 더 커지고 95% 신뢰구간 포함률은 0.95보다 더욱 작아짐을 알 수 있었다.

구간 중도절단된 자료는 지속적인 추적관찰이 이루어지는 임상연구에서 흔하게 접할 수 있다. 이러한 자료가 관측되었을 때 많은 임상연구에서는 비교적 간단하게 구간 중도절단된 자료를 처리할 수 있다는 이유로 중간값 대체법이나 우측값 대체법을 사용하고 있다. 하지만, 이러한 두 대체법은 편향된 추정을 하게 되어 잘못된 결과를 도출할 가능성이 커지게 된다. 이와 같이 단일 대체법이 갖고 있는 단점을 극복할 수 있는 방법으로 다중 대체법 [12]을 대안으로 제시할 수 있다. 이 방법은 단일 대체법과 달리 구간 내 사건발생의 확률을 갖는 값들 중에서 확률적으로 대체하고 또한 여러 번 반복적으로 대체함으로써 단일 대체법이 갖는 문제점을 보완한 방법이다. 다만 본 논문에서는 'coxph' 함수로는 다중 대체법을 구현할 수가 없어서 다중 대체법에 비해 단순하지만 'coxph' 함수로 구현가능한 단일 대체법만을 적용하였다. 그러나 앞으로 단일 대체법보다 추정량의 Bias나 평균제곱오차 측면에서 더 우수할 것으로 예상되는 다중 대체법을 구현하여 단일 대체법과 다중 대체법, 그리고 'intcox' 방법을 서로 비교하는 연구가 필요하다고 생각된다.

아직까지 대부분의 통계소프트웨어들은 본 논문에서 다룬 Cox 비례위험모형의 회귀계수에 대한 추정값을 제공하지만 추정량의 표준오차는 즉,

신뢰구간은 제시하지 않고 있다. 앞으로 구간 중도절단된 자료를 분석할 때에는 구간 중도절단된 자료의 특성을 잘 파악하고 적절한 분석방법을 선택하여야 함을 본 연구는 시사한다.

## References

- [1] Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society* 1976;38(3):290-295.
- [2] Groeneboom P, Wellner JA. Information bounds and nonparametric maximum likelihood estimation (Vol. 19). New York: Springer Science & Business Media; 1992.
- [3] Wellner JA, Zhan Y. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association* 1997;92(439):945-959.
- [4] Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics* 1986;42(4):845-854.
- [5] Pan W. Extending the iterative convex minorant algorithm to the Cox model for interval-censored data. *Journal of Computational and Graphical Statistics* 1999;8(1):109-120.
- [6] Finkelstein DM, Wolfe RA. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* 1985;41(4):933-945.
- [7] Lindsey JC, Ryan LM. Tutorial in biostatistics methods for interval censored data. *Statistics in Medicine* 1998;17(2):219-238.
- [8] Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken: John Wiley & Sons; 2004.
- [9] Cox DR. Partial likelihood. *Biometrika* 1975;62(2):269-276.
- [10] Henschel V, Heiss C, Mansmann U. intcox: Compendium to apply the iterative convex minorant algorithm to interval censored event data. 2009.
- [11] Kim J. Effect of thyroidectomy on the time to recovery of parathyroid hormone of thyroid carcinoma patients. *Journal of the Korean Data Analysis Society* 2011;13(5):2301-2314 (Korean).
- [12] Pan W. A multiple imputation approach to Cox regression with interval censored data. *Biometrics* 2000;56(1):199-203.